A METHODOLOGY FOR DESIGNING A RECOMMENDER SYSTEM BASED ON CUSTOMER PREFERENCES

Inès Jomaa, Emilie Poirson, Catherine Da Cunha and Jean-François Petiot
LUNAM, Ecole Centrale de Nantes, IRCCyN UMR CNRS 6597, France

ABSTRACT
This paper presents a contribution to design an online preference based system. The objective of the system is to assist a customer in the products selection process. Current e-commerce recommendation systems assist customers in this process. Nevertheless, quality of the recommendations produced remains a real challenge. There are products that are by mistake recommended to customers and inversely. This paper focuses on these quality and relevance of recommendations. More than product objective characteristics, the customer's choice is also based on his/her perceptive expectations. Therefore, to be considered as relevant, products recommendations must reach customer's expectations and particularly perceptive ones, sometimes spontaneously, without specific request. Collaborative filtering and neighbourhood formation are the main tools used. The cluster of "perceptive" neighbours containing the active customer share common perceptive preferences and can guide the propositions. The application case is the comic. The aim is to propose to a customer a "good" product. A test procedure enabling the validation of this algorithm is to be set.

Keywords: Recommender Systems, Collaborative Filtering, Emotions, Semantics, preference

1 INTRODUCTION
Appeared in the mid 1990s, recommender systems are used by electronic commerce websites. These websites offer a huge number of products and services for sale, therefore choosing the right product is not an easy task for a website visitor. Recommender systems have emerged in order to help people to choose products or services that best meet their needs and preferences. These systems are based on a variety of filtering techniques such as collaborative [1], [2], content-based [3], [4], demographic filtering [5], [6] and other techniques [7], [8].
To recommend products to an active customer, most of recommendation algorithms focus on finding similar customers from similar products they purchased or rated. In fact, collaborative filtering [9], [10] is one of the earliest and most promising recommender technique, it makes recommendations by matching people with the same interest. It predicts products a customer would like, based on the opinions (evaluations or ratings) of other people for these products. In order to identify people who share the same interests, similarities between customers are calculated by determining the proximity between them. The predicted score attributed by the active customer to unseen products is determined using the evaluation scores of his/her neighbourhood.
Collaborative filtering offers to customers the possibility to discover other various domains. In fact, diversity and novelty characterize the recommendations produced. Collaborative filtering also works well for complex objects such as movies and music [7]. However, collaborative filtering suffers from several limitations such as the "Sparsity problem" [11], [8], the "Gray sheep" [11] and the "quality of recommendations for the users" [12]. Indeed, sparsity problem occurs when there are not enough common products rated by customers. For example, if the number of customers is small relative to the large number of products, the matrices containing the customers' ratings of products could be sparse, thus it is hard to find correlations between customers and consequently to find neighbourhoods. "Gray sheep" problem occurs in a small community of customers, when there are customers having no close neighbours and whose ratings don't match any group. Consequently, they get poor recommendations.
The collaborative filtering challenge addressed in this paper concerns the quality of recommendations. In fact, for several recommender systems, there are products that are not recommended to one customer though he/she would appreciate them and others recommended to him/her though he/she does not appreciate them [12]. Thus, recommendations produced are not always relevant. For

example, let's imagine that Paul and Edouard like detective stories; but their choice of liking cookery books are different as Paul like them but Edouard don't. Collaborative filtering could recommend cookery books to Edouard, even though these recommendations are not relevant for him.
The problem of relevance of recommendations produced by collaborative filtering algorithm is particularly addressed in this paper. The aim is to propose one possible extension of collaborative filtering based on emotional, sensory and semantic product characteristics. Our proposed algorithm includes an improvement of understanding perceptive customer needs and can improve the finding of potentially interesting products. In fact, considering perceptive - particularly emotional- similarities between customers and explaining the overall customer preference basing on the product perceptive evaluation represent one contribution. Products obtained meet customers' perceptive expectations.
In order to validate the algorithm, a test procedure is described. This procedure simulates the customers' behaviours, removes some products and checks if the algorithm recommends the right ones.
The product chosen within the framework of our project is the comic. The comic market is growing and the attendance to various festivals devoted to comics is very important.
The paper is organized as follows. In Section 2, we present the process of determination of the comic's characteristics named comic's attributes on which the recommender system is based. Section 3 describes the system framework. In Section 4, we present the algorithm that will be implemented. Finally, in section 5, we present a testing procedure in order to validate the algorithm.

2 DETERMINATION PROCESS OF THE PRODUCT ATTRIBUTES

Two categories of product attributes are considered here: *objective* and *perceptive* attributes. Characteristics of the product that are independent of the user and that will be used to define the product sheet are called *Objective* attributes (e.g. number of pages, author, …). *Perceptive* attributes are divided into two categories: *emotional* category and *sensory* and *semantic* category. We name *emotional* attributes those related to the affective state of the user, pleasure and displeasure. *Sensory* and *semantic* attributes are all subjective attributes that don't belong to the *emotional* category.

2.1 Determination of the objective attributes

Objective attributes will be used to define the comic's sheet. They will also be used by the customer in his/her search of comics. In order to determine them, two comics' experts (a bookseller and comic association journalist, a comic bookseller) were questioned about relevant comic objective attributes. A comic website was also referred. A list of objective attributes was obtained. The price is considered important in the definition of a comic since it represents generally a criterion for buying a product for customer. According to the different sources of information, the list of the 11 most relevant objective attributes was finally defined:
Title of series (or Title of album if there is only one album); volume number – title of volume; screenwriter(s); cartoonist; colorist; publisher; collection; number of pages; edition release date ; summary (either one for the series or one for the volume, but often, there is one summary per volume), price

2.2 Determination of the emotional attributes

A summary list of emotions described in the literature was presented to ten comics' fans [13], [14] [15], [16], [17]. They were asked to identify the main representative emotions provoked by reading a comic.
The 10 most frequently cited emotions were the following: *joy, surprise, pleasure, curiosity, fascination, relaxation, boredom, satisfaction, awaiting,* and *calm*. This list of terms was retained for future use.

2.3 Determination of the sensory and semantic attributes

Terms related to sensory and semantic comics' characteristics were collected by two different ways: from readers' comments from two comics' websites and by questioning comics' fans. We also refer to McCloud [18].
The most occurring terms were considered. Among the identified terms, the term "emotional" was got. This term was considered encompassing all emotions identified in the previous section; therefore it was omitted from the final list of sensory and semantic attributes. The term "calm" was also omitted

since it was considered as an emotional term. Finally, the following list of 10 comic's sensory and semantic attributes was kept:

Gripping, nice, classic, clear, imaginary, original, artistic, funny, dynamic, of quality.

3 PREFERENCE BASED SYSTEM FRAMEWORK

The purpose of our system is to assist customers (users of the system) in their choice of an appropriate comic; this comic is assumed to meet their expectations. The system uses three data bases (figure 1), namely:

- **Customers' database**: stores personal information of each customer. The customer's expectations (that is to say the expected objective attributes of the comic) will also be stored. In order to be registered, each customer has to fill this information form. Data in customer's form (figure1) represents the values of his/her personal information (name, age…) and the expected objective attributes (collection, publisher…), if any. An attribute can have multiple values (ex. One customer likes more than one comics' publisher).
- **Products' database**: stores the value of the objective attributes for each comic (comics' sheets, figure1). This data base is updated once a new comic appears in the market.
- **Preferences database**: stores the ratings of each comic by the customer. To that end, a product evaluation form is provided (figure1). Global rating e_{ij} represents the overall evaluation of a comic *i* by a customer *j*. It reflects customer's preference regarding the comic that he/she had read and it is valued on a 10 point-scale. The different perceptive comic's attributes are also evaluated on a Semantic Differential Scale [19]. Here, we consider a 7 point-scale (Product evaluation form, figure1).

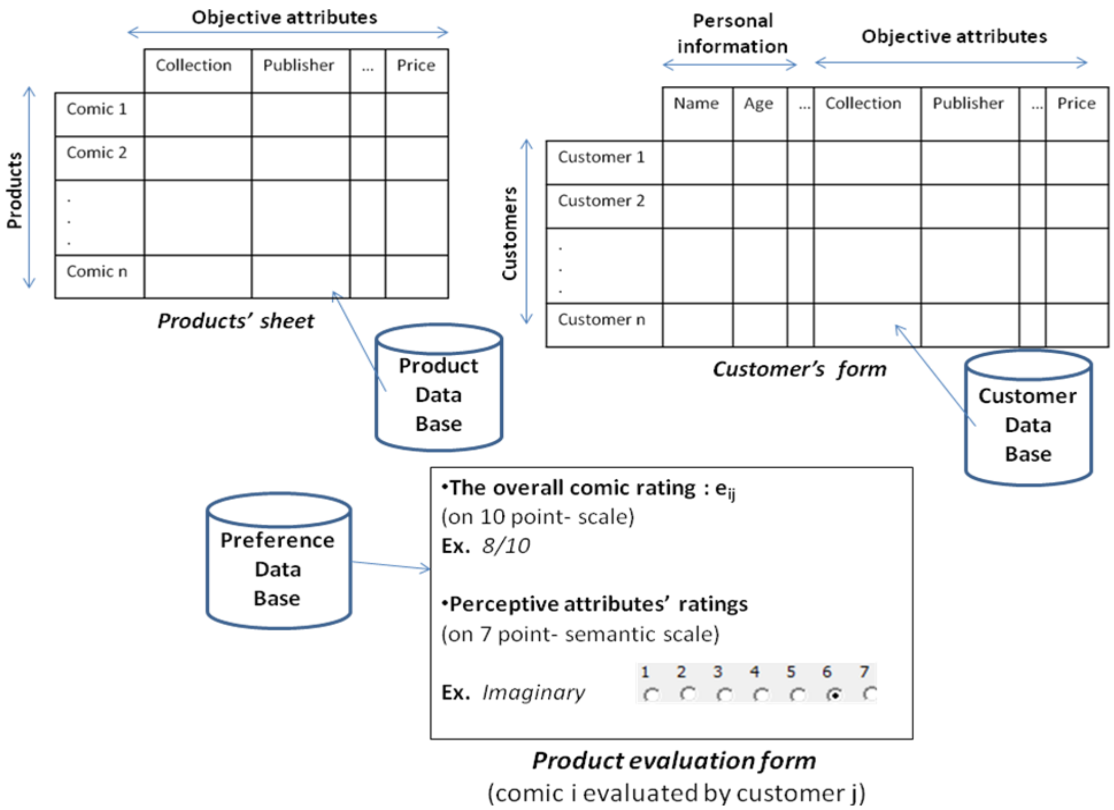

Figure 1. Illustration of the databases' structure

4 BASIC STAGES OF THE PROPOSED FILTERING ALGORITHM

The system use case considered is: "The system spontaneously recommends products to registered customer". Recommendations are based on customers' comics' past evaluations.

Figure 2 illustrates by an example the different stages of the algorithm applied to an active customer called "Paul". The algorithm proceeds in 3 stages:

- Stage 1: determination of the " general neighbourhood" (Cluster 1) of the customer, based on the global evaluations e_{ij} of the products.
- Stage 2: reduction of this neighbourhood to people who evaluate similarly the comics, for the same reasons: $Cluster_{selection}$. The focus is on the reasons of the evaluation (evaluation of the objective and perceptive criteria which explain the global evaluation of the products).
- Stage 3: selection of the best proposition for the customer by filtering.

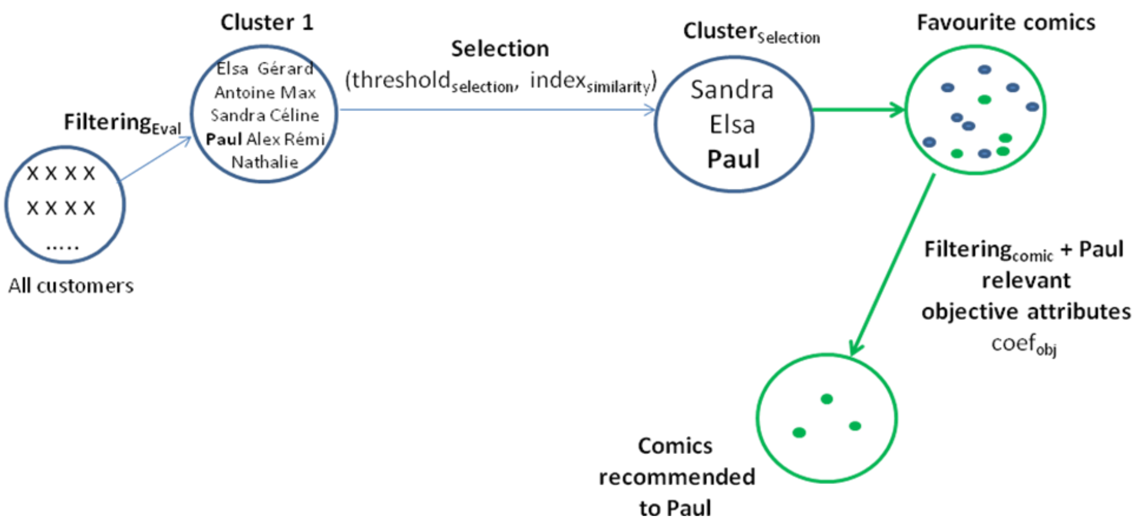

Figure 2. Stages of the proposed algorithm

The next section presents in details each stage of the algorithm. "Paul" will be used as our central customer.

4.1 Stage 1: Filtering$_{Eval}$

As presented in section 3, the customer "Paul" evaluates each comic he/she read with:

- a global evaluation e_{ij}, (*i* is the index of the comic read by Paul, and *j*=Paul)
- evaluations of the perceptive attributes

This stage consists in applying a filtering on all customers of the system on the basis of the global ratings (or the overall evaluation) assigned to each comic.
Let *n* be the number of comics in the system and *c* the number of customers (figure 3).

	Customer 1	Customer 2	Customer j ...	Customer c
Comic 1	$e_{1,1}$	....		
Comic 2	...			
.				
.				
Comic i			$e_{i,j}$	
.				
.				
Comic n	...			$e_{n,c}$

Figure 3. Illustration of the overall ratings of comics by customers

A measurement of the proximity between two customers is required. This measurement is achieved by the method of the cosine measure [20]. In fact, each customer A of the matrix (figure 3) can be represented by a vector $\vec{A}$. The Cosine method calculates the cosine of the angle between two given customers A and B. It represents the similarity between them. A similarity matrix (*c* X *c*) of all customers is obtained. Then, "Center-based" method is used. It consists in selecting the k nearest

neighbours to a particular customer; k represents the maximum number of neighbours or threshold that has to be set.
In order to form Paul's neighbourhood, we assume that customers compared with Paul must have evaluated at least two comics in common with him. The proximity between Paul and one customer *x* is the similarity coefficient $sim_{Paul,x}$, determined by the cosine similarity [1]:

$$sim_{Paul,x} = \frac{\overrightarrow{Paul}.\vec{x}}{\left\|\overrightarrow{Paul}\right\|_2 \left\|\vec{x}\right\|_2}.$$

$\vec{x}$ represents the vector of marks given by the customer x to the evaluated comics.

$\left\|\vec{x}\right\|_2$ represents the Euclidian norm of $\vec{x}$

A maximum number *MaxN* of neighbours is defined, in order to limit the size of Paul's neighborhood. Cluster 1, including the *MaxN* Paul's nearest customers, basing on global evaluations is formed.

4.2 Stage 2: Selection ($threshold_{selection}$, $index_{similarity}$)

The first stage giving Paul's nearest customers is only based on global evaluations. As explained in the introduction, the problem is not only to get customers having same evaluations, but, also customers who have the same reasons of evaluations (the same perceptive evaluations). Thus, the second filtering is based on evaluations of the comics' perceptual attributes, in order to form $Cluster_{selection}$ (figure 2).
Let *p* be the number of comics evaluated by Paul. For each comic *i* evaluated by Paul, a perceptive proximity between Paul and one customer *x* (belonging to cluster1) is determined. This proximity is the similarity between Paul and the customer, based on the perceptive evaluations of the considered comic, called Sim (Paul, $x)_i$.
Sim (Paul, $x)_i$ is determined as follows: for each comic evaluated by Paul, a matrix is associated: perceptive attributes are in rows and customers are in columns (figure 4). The coefficient of this matrix ($e_{attribute,\ customer}$) is the evaluation of a comic perceptive attributes by a given customer. The evaluation of each attribute is rated on 7 point-semantic scale (see "Product evaluation form", figure 1). Each customer of the matrix can be represented as a vector of $e_{attribute,\ customerx}$. In the same way as the previous section, cosine similarity method is used to determine Sim (Paul, $x)_i$.

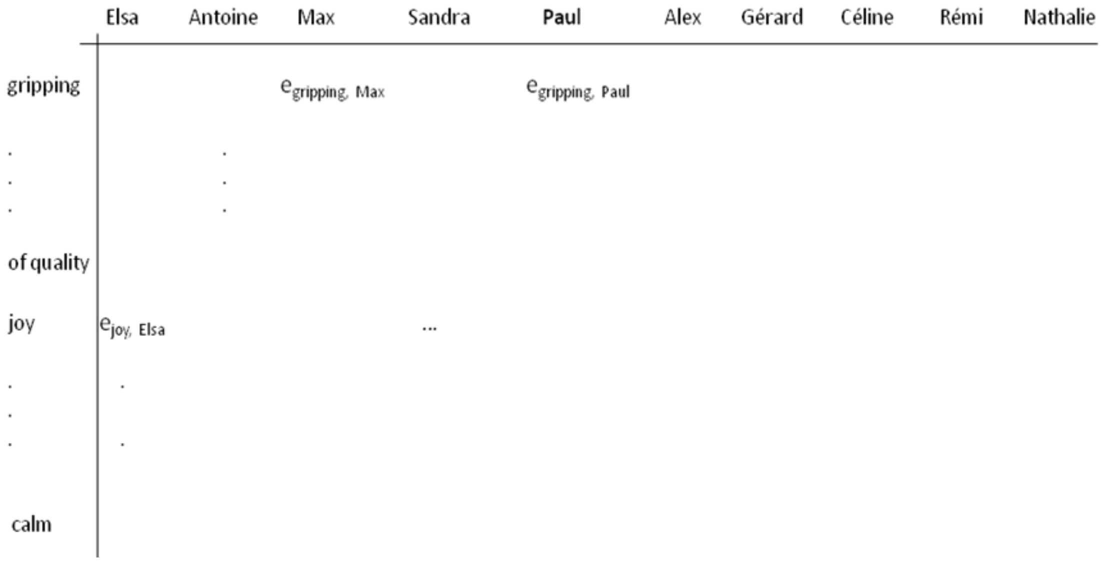

Figure 4. Matrix of perceptive evaluation of comic by cluster1's customers

Let *p'* be the number of comics evaluated by Paul and customer *x* at the same time ($p' \leq p$). From the matrix presented in figure 4, it is possible to determine the overall similarity coefficient between Paul and Customer *x*. The overall similarity coefficient is given by the following formula:

$$\text{index}_{\text{similarity}} = \frac{1}{p}\sum_{i=1}^{p'} Sim(Paul, x)i$$

$\text{index}_{\text{similarity}}$ is determined for each Paul's neighbour. Cluster1 is sorted by "$\text{index}_{\text{similarity}}$". A threshold ($\text{threshold}_{\text{Selection}}$) is set. It corresponds to the size of Paul's perceptive neighborhood. This selection leads to $\text{Cluster}_{\text{selection}}$ which includes neighbours who share Paul's perception of emotional, sensory and semantic attributes.

We consider here that a good (pleasant) comic has an evaluation higher than 7/10. Favourite comics of each customer who belongs to $\text{cluster}_{\text{selection}}$ are determined. The set of the favorite comics includes those having an average evaluation above 7/10 and not already read by Paul.

4.3 Stage 3: Comics recommended to Paul

The set of the favourite comics determined in stage 2 must be reduced in correlation to Paul's relevant objective attributes (i.e. his ideal comic given in the customer's form). This step is called $\text{Filtering}_{\text{comic}}$. The used method is based on a threshold coef_{obj}. This threshold defines the maximum distance between Paul's relevant objective attributes and comic's objective attributes. In order to compute this distance, a dummy coding (binary matrix) is used (figure 5).

	comic 1	comic 2	comic 3	...	comic c
Objective attribute 1	*1*	*0*	*1*	...	*1*
Objective attribute 2	*0*	*1*	*1*	...	*1*
.	.	.	.	...	.
.	.	.	.	...	.
.	.	.	.	...	.
Objective attribute 11	*1*	*0*	*0*	...	*1*

Figure 5. Dummy coding of the objective attributes- Matrix showing the absence or presence of Paul's relevant objective attributes in the favourite comics

If one favourite comic meets one of the Paul's relevant objective attributes (or expectations), "1" is inserted in the corresponding row. Otherwise, "0" is placed. For each comic (column), the number of "1" is counted and divided by the number of objective attributes. The result represents the distance d_j between Paul's ideal comic and a favourite comic *j* (*j*=1…c):

$$d_j = \frac{\sum_{i=1}^{nbObj} b_{ij}}{nbObj}$$

b_{ij} =1 or 0; nbObj the number of objective attributes. (Here, nbObj=11)

d_j is compared with a threshold coef_{obj}.

The final result is a set including the most relevant products, taking into account Paul's objective and perceptive expectations. This set is recommended to Paul.

4.5 Summary of the algorithm thresholds

In the proposed algorithm, three parameters have to be set. The size of the active user's neighbourhood *MaxN* which represents the number of the nearest neighbours has to be defined. This parameter allows the formation of the cluster1.

Moreover, in order to select customers who share perceptive expectations, another parameter $\text{threshold}_{\text{Selection}}$ has to be determined. It represents the size of the active user's perceptive neighbourhood. Finally, a parameter $coef_{obj}$ that defines the relevant recommended comics - basing on the active user's objective expectations- has to be set.

5 TEST PROCEDURE

In order to assess the performances of the proposed algorithm, a test procedure has to be implemented. It consists in simulating the customers' behaviour. The principle of the procedure is to define a priori an explicit preference function for the customers, which estimates the overall evaluation e_{ij} of the comic according to the assessment of the perceptive attributes. A set of customers and products will be simulated to fill the different databases.

Next, the principle of the test will be to remove appreciated products to the "simulated" customers, and to check if the system recommends the correctly ones (procedure similar to a cross-validation).

6 THEORETICAL CONTRIBUTIONS

Considering the customers perceptive similarities is one of the theoretical contributions of our study. Determining the customer's perceptive neighbourhood based on the perceptive evaluations of the products could be better than using the traditional approach considering an overall evaluation. In fact, to show that the proposed algorithm is better than the traditional one, it has to be tested.

Another contribution of this paper is the interpretation of the overall product evaluation in terms of customer objective expectations and product perceptive evaluations.

7 VALIDITY OF THE APPROACH WITH OTHER PRODUCT TYPES

The algorithm upstream phase that consists in determining the product attributes depends on the product type. In order to apply the method defined above to different product types such as food, clothes..., this phase should be modified. More precisely, the algorithm inputs, which are the objective and perceptive attributes, should be examined since they are determined differently from one product type to another.

In fact, objective attributes are different according to the product type; by definition, they are directly related to the product characteristics.

Sensory and semantic perceptive attributes as presented in this work are determined from information related to the comics. Nevertheless, the first list of emotional attributes was derived from the literature where they are presented regardless of the product type. Consequently, this list can be a starting point for different product types. In fact, it describes emotions provided by a product and emotional attributes could be transposed to other product types. Nevertheless, it may be more interesting to define a relevant list choosing the emotional attributes that are the most generic ones.

Similarly, suggesting a generic methodology applicable to different product types could be more efficient. This methodology allows the determination of the product relevant attributes and may improve the recommendation algorithm quality. The objective is then to establish a list of relevant attributes in order to get a better idea of the customer preference and realize an intelligent filtering.

8 CONCLUSION, DISCUSSION AND PERSPECTIVES

The purpose of this project is to design a preference based system that helps customers in their purchase decision. Different use cases are considered. In this paper, the use case that consists in making spontaneous recommendations to the customers of the system is chosen. To that end, an algorithm, which recommends a set of products (in our case comics), is suggested. This algorithm is based on the traditional collaborative filtering, selection techniques and clusters constitution based on the similarity between the customers. Clusters gather the customers who share similar preferences toward products. The algorithm takes into account the products' evaluations, and a refining of perceptive (emotional, sensory and semantic) data; here is our contribution. In fact, refining improves the relevance of recommendations. Recommended products answer customer objective and perceptive expectations. However, "Sparsity" and "Gray sheep" problems, which represent limitations in collaborative filtering persist in the suggested algorithm. Moreover, threshold parameters remain to be set.

It is also interesting to compare the proposed algorithm to the traditional one; the algorithm which doesn't take into account the product perceptive evaluations.

The proposed algorithm can be applied with other product types. However, the inputs that are objective and perceptive attributes may depend on the product type. The process of determination of these attributes may also depend on the product type. A methodology for determining the accurate attributes could be suggested. It is not discussed in this paper. Suggesting a methodology applicable to other product types is an interesting contribution.

As perspectives of this work, different types of perceptive data can be separated to distinguish different evaluation methods and to obtain different corresponding customers' clusters. The test and the validation of the proposed algorithm remain to be done. The test procedure is under way. It will validate our work and allows the determination of the various parameters that have to be set.

Furthermore, other use cases of the system remain to be released. These use cases represent other possible interactions between the customers and the system, namely: a customer enters his/her personal information, and his/her expectations in the system; a customer searches comics depending on different criterion. Another use case in relation to the storing and the update of the products in the product data base has to be realized.

REFERENCES

[1] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. GroupLens:An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of CSCW '94,* Chapel Hill, NC.

[2] Hill, W., Stead, L., Rosenstein, M., and Furnas, G. Recommending and Evaluating Choices in a Virtual Community of Use. In *Proceedings of CHI'95.*

[3] Lieberman, H. *Letizia:* An Agent that Assists Web Browsing. In *Proceedings of the IJCAI'95,* 924–929.

[4] Pazzani, M., Muramatsu, J. & Billsus, D. Syskill & Webert: Identifying Interesting Web Sites. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence,*1996, 54–61.

[5] Krulwich, B. LifeStyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine* 18(2): 1997, 37–45.

[6] Pazzani, M. A Framework for Collaborative, Content-Based and Demographic Filtering. Department of Information and Computer Science. *University of California, Irvine. Irvine, CA 92697,* 1999

[7] Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), 2002, 331-370.

[8] Montaner M., López B., De La Rosa J.-L. A Taxonomy of Recommender Agents on the Internet, *Artificial Intelligence Review*, 2003, vol. 19, p. 285-330 (Kluwer Publishers).

[9] Shardanand, U., & Maes, P. Social information filtering: algorithms for automating "Word of Mouth". In *Proceedings of the SIGCHI conference on Human factors in computing systems,* Denver, 1995, Colorado, United States (ACM Press/Addison-Wesley Publishing Co).

[10] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 1997, 40(3), pp. 77-87.

[11] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M. Combining Content-based and collaborative filters in an online newspaper. *June 22, 1999*

[12] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. Analysis of Recommendation Algorithms for E-Commerce. *2nd ACM Conference on Electronic Commerce*, 2000, pp. 158-167

[13] Russell, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 1980, 39, 1161-1178.

[14] Russell, J. A. Core affect and the psychological construction of emotion. *Psychological Review,* 2003, Vol 110, No 1, 145-172

[15] Plutchik, R. *Emotion: a psychoevolutionary synthesis,* 1980 (New York: Harper & Row).

[16] Desmet, P.M.A. and Hekkert, P. Framework of Product Experience. *International Journal of Design,* 2007, Vol1 No 1 2007

[17] Fox, E. *Emotion Science*, 2008 (Palgrave Macmillan).

[18] McCloud, S. Making Comics, 2006 (Harper)

[19] Osgood, C.E., Suci, C.J., Tannenbaum, P.H. *The Measurement of Meaning*. University of Illinois Press, Urbana, 1957, pp. 76}124.

[20] Breese, John S., Heckerman, D. and Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering, 1998 *Technical Report MSR-TR-98-12*

Contact: Inès Jomaa, Jean-François Petiot
LUNAM, Ecole Centrale de Nantes, Institut de Recherche en Communications et Cybernétique de Nantes (UMR CNRS 6597)
1, rue de la Noë, BP 92101, 44321 NANTES France
Tel: Int +33 (0)2 40 37 69 22
Fax: Int +33 (0)2 40 37 69 30
Email: ines.jomaa@irccyn.ec-nantes.fr; Jean-Francois.Petiot@irccyn.ec-nantes.fr
URL: http://www.irccyn.ec-nantes.fr/~petiot/

Inès Jomaa is a PhD Student in IRCCyN (Institut de Recherche en Communications et Cybernétique de Nantes, UMR CNRS 6597), Ecole Centrale de Nantes. She received her Master's degree on Industrial Engineering from Ecole Centrale Paris. Her research interests include customer centred design, especially customer perception-oriented product design.

Emilie Poirson is Associate Professor of engineering design at Ecole Centrale de Nantes.
Her point of interest, both in teaching and research concern the taking into account of the perceptions of users in product design. She particularly focuses her work on the consideration of human variability in design, including the study of disability for product design for all (Universal design).

Catherine da Cunha is Associate Professor in Industrial Engineering at the École Centrale de Nantes, France. She received her Ph.D from the INP Grenoble, France in 2004. Her main interests are in the area of information use within the firm, in the production as well as in the organizational field. She is particularly interested in the potential of partial or unmature information for mass customization implementation.

Jean-François Petiot is Professor of Engineering Design at the Ecole Centrale de Nantes. He teaches and researches in customer oriented design and user centered design. He is interested by many aspects of design, in particular user's perceptions, subjective assessments, sensory analysis, preference mapping, psychoacoustics. He has a deep interest for applications in acoustics and in the domain of the perceived quality.